第65卷 第3期 天 文 学 报 Vol. 65 No. 3
2024年5月 ACTA ASTRONOMICA SINICA May, 2024

doi: 10.15940/j.cnki.0001-5245.2024.03.010

基于FPN-ViT的星系形态分类研究*

曹 婕[1] 许婷婷[1†] 邓雨禾[1] 李广平[1] 高献军[1] 杨明存[1] 刘执靖[1]
周卫红[1,2‡]

(1 云南民族大学数学与计算机科学学院 昆明 650504)
(2 中国科学院天体结构与演化重点实验室 昆明 650011)

摘要 随着人工智能技术的发展, 利用深度学习方法进行星系形态分类研究取得了较大进展, 但在分类精度、自动化及其星系的空间特征表示上仍然存在不足之处. Vision Transformer (ViT)模型目前在星系形态分类上具有较好的鲁棒性, 但是在处理多尺度图像时存在一定的局限性, 因此提出将特征金字塔(Feature Pyramid Networks, FPN)引入ViT模型(FPN-ViT)中进行星系形态的分类研究中. 结果表明: 基于FPN-ViT模型进行星系形态分类的平均准确率、精确率、召回率以及F1分数等各项评估指标均在95%以上, 与传统的ViT模型相比各项指标均有一定程度的提升. 同时, 在原始星系图像中加入不同程度的高斯噪声和椒盐噪声, 验证FPN-ViT模型对低信噪比数据也能获得较好的分类性能. 此外, 为了对模型进行综合评估, 采用t分布随机邻接嵌入(t-distributed Stohastic Neighbor Embedding, t-SNE)算法对分类结果进行了可视化分析, 能够更加直接地看出FPN-ViT模型对于星系形态分类的效果. 因此, 将FPN网络应用于ViT模型对星系形态的分类研究中是一种全新尝试, 对后续研究具有重要意义.

关键词 方法: 数据分析, 技术: 图像处理, 星系: 普通

中图分类号: P152; **文献标识码**: A

1 引言

星系形态分类是天文学中用于描述和分类不同星系外观和结构的系统. 通过星系的不同形态进行分类, 其目的是通过考察星系的形状、大小、密度和结构等特征, 来深入了解星系形成和演化的物理过程并揭示宇宙的大尺度结构. 通过观察大量的星系, 并将它们分为不同的形态类别, 我们可以了解星系形成的不同机制和演化的不同路径[1]. 例如, 早期的星系形态分类系统中包含了旋涡星系和椭球星系这两个主要类别. 这些类别的存在揭示了星系形成的两种不同机制: 旋涡星系通常是由气体和尘埃云的坍缩和自转形成的, 而椭球星系则很可能是通过星系碰撞和合并形成的. 与此同时, 星系形态分类还与星系的物理性质和环境相关联. 例如, 在物理性质方面: 研究发现旋涡星系通常富含年轻恒星和星际物质, 而椭球星系则倾向于包含年老的恒星和较少的星际物质. 在环境关联方面: 研究表明在星系团等高密度环境中, 星系之间的相互作用和相互影响更加显著. 这可能导致星系形态的变化, 例如, 星系之间的相互作用可能促使星系转变为椭球形态. 因此, 星系形态分类可以提供关于星

2023-05-22收到原稿, 2023-08-02收到修改稿

*国家自然科学基金项目(61561053)、云南省教育厅科学研究基金项目(2023J0624)资助

†xutingting@cnlab.net

‡ynzwh@163.com

系所处环境的重要线索[2–3]. 这种关联有助于我们理解星系内部物质的分布和演化过程. 最后, 通过研究不同类别的星系在宇宙中的分布和聚集程度, 我们可以揭示宇宙的网状结构、星系团和超星系团等大尺度组织形式[4]. 这对于理解宇宙的演化等基本物理问题具有重要意义. 因此, 将星系按照形态特征进行准确分类是后续数据分析和挖掘的基础.

星系的形态可以根据不同的分类标准进行划分. 其中, 哈勃于1926年提出的哈勃序列(Hubble sequence)是最著名的早期星系形态分类标准之一. 哈勃序列与中性氢的质量、星系的积分颜色、星系光度和环境等物理参数密切相关, 至今仍具有重要的参考价值. 2007年推出的星系动物园项目(Galaxy Zoo, GZ[5])采用的星系形态分类标准就是基于哈勃序列, 该项目通过广泛的众包志愿者参与, 以哈勃序列为基础对星系的形态进行分类, 志愿者们根据星系的结构特征、旋臂存在与否、条纹形状等属性, 将星系分为不同的类别. 哈勃序列作为一种经典的星系形态分类标准, 对于研究星系的演化和形成过程仍然具有重要的意义.

近年来, 天文观测设备的巡天深度和探测效率不断提升, 斯隆数字巡天(Sloan Digital Sky Survey, SDSS[6])、郭守敬望远镜(The Large Sky Area Multi-Object Fibre Spectroscopic Telescope, LAMOST[7])等巡天项目和詹姆斯·韦伯空间望远镜(James Webb Space Telescope, JWST[8])等红外线太空望远镜观测产生了海量的星系光谱数据和图像数据, 因此迫切需要寻求更加自动化和智能化的分类方法以满足大规模的星系图像数据处理需求. 随着深度学习技术的不断发展, 深度学习相关算法被广泛应用于天文领域, 其中基于深度学习算法的星系形态分类研究就是研究热点之一. Zhu等[9]提出了基于深度残差网络(Residual network, ResNet)的改进模型, 模型名为ResNet-26, 即具有26层(25个卷积层和1个全连接层), 该模型实现了对星系形态特征的自动提取、识别和分类. 实验结果表明: ResNet-26模型的分类精度达到了95.12%, 与其他流行的卷积神经网络(Convolutional Neural Network, CNN)模型相比具有更好的分类性能. 艾霖嫔等[10]在星系形态分类研究中应用了EfficientNet模型, 实验结果显示: 基于EfficientNet-B5模型进行分类的各项评价指标均超过96.6%, 与使用残差网络中效果较好的ResNet-26模型相比, EfficientNet模型在分类结果上取得了显著的提升. Wei等[11]提出了一种基于对比学习的方法, 针对星系图像语义信息少、轮廓占主导的特点, 在特征提取层采用视觉变换器和卷积网络相结合, 通过融合多层次特征提供丰富的语义表示, 并在Galaxy Zoo 2, SDSS发布的第17个版本数据以及Galaxy Zoo DECaLS 3个数据集上训练和测试了这个方法, 测试集准确率分别达到94.7%、96.5%和89.9%. He等[12]基于具有50层(49个卷积层和1个全连接层)的ResNet-50网络结构设计了一个多通道深度残差网络框架ResNet-Core, 分别针对光谱图像和星系图像的特点, 通过加入卷积核方差控制技术提取轮廓和细节特征, 有效提高了平均精度, 超过了当时最高性能的ResNet-50, 结果表明ResNet-Core模型具有更好的分类性能和更好的鲁棒性. Hui等[13]提出将稠密卷积网络(Densely connected convolutional Networks, DenseNet)算法应用于星系形态分类中, 实验结果显示, 使用具有121层(120个卷积层和1个全连接层)的DenseNet-121模型得到的准确度为91.79%, 也就是在3044张测试图像中, 能够准确分类出2794张星系图像; 另外, 模型的精确度为79.92%、召回率为73.20%、F1分数(F1-Score)为75.48%. Li等[14]提出了一个多尺度卷积胶囊网络(Multi-Scale Convolution Capsule Network, MS-CCN)模型进行星系形态分类研究, 该模型通过使用多分支结构来提取星系图像的多尺度隐藏特征, 并在Galaxy Zoo 2这一数据集上进行训练和测试, 实验结果表明该模型在宏观平均下达到97%的准确率、96%的精确率、98%的召回率和97%的F1分数.

2021年开始, 深度学习中的Transformer模型通过引入自注意力机制, 实现了对序列数据的全局上下文建模, 并且在自然语言处理(Natural Language Processing, NLP)中取得了巨大的成功. 同时Google团队开发出了一种新的图像分类架构, 称为Vision Transformer (ViT)[15]. ViT模型发布至

今已经被广泛运用于各个领域的分类任务, Ghe-flati等[16]将ViT模型应用于医学领域, 对乳腺超声图像进行分类, 结果表明ViT模型对于乳腺超声图像的分类效果比CNN模型更好. Gao等[17]用ViT模型参与了人工智能医学图像分析COVID-19诊断竞赛挑战, 根据计算机断层扫描(Computed Tomography, CT)技术得到的肺部CT图像将新冠肺炎与非新冠肺炎进行分类, ViT模型的结果优于同期参赛的DenseNet模型, F1分数为0.76. Tanzi等[18]采用ViT体系结构对不同骨折类型图像进行分类, 并与经典CNN和由连续CNN组成的多级结构进行了比较, 结果显示ViT模型能够正确预测83%的测试图像, 性能优于CNN模型.

在Vision Transformer发布之后, 许多研究者对模型进行了改进. 例如: Chu等[19]提出一种条件位置编码视觉Transformer (Conditional Position encodings Visual Transformer, CPVT)结构, 使用条件位置编码(Conditional Position Encodings, CPE)代替ViT中的预定义位置嵌入, 使Transformers能够处理任意大小的图像且无需插值. Han等[20]提出Transformer-iN-Transformer (TNT)模型, 该模型利用处理图像块嵌入的外部Transformer模块和对像素嵌入之间的关系进行建模的内部Transformer模块来对补丁级和像素级表示进行建模. Yuan等[21]提出了Tokens-To-Token (T2T)模型, 主要通过将滑动窗口内的多个token连接成一个token来改进ViT模型. Wang等[22]提出了Pyramid Vision Transformer (PVT)模型, 它为Transformer采用了多级设计(没有卷积), 类似于CNN中的多尺度, 有利于密集预测任务. Wu等[23]提出卷积视觉Transformer (Convolutional vision Transformer, CvT), 将卷积引入到ViT模型, 以提高ViT模型的性能. CvT模型在公开数据集ImageNet-1k上获得了87.7%的Top-1准确率(表示模型的第1个预测是否与实际标签相符的比例, 是最常用的分类性能指标), 超过了ViT模型在该数据集上76%的准确率.

基于上述对ViT模型的改进方法, 本文提出将特征金字塔网络(Feature Pyramid Networks, FPN)引入ViT模型, 以提高模型的性能. 本文的组织结构如下: 首先讨论了FPN和传统的ViT网络结构, 并在第2节介绍了将FPN引入ViT之后组成的FPN-ViT网络架构基本框架和原理; 在第3节中, 我们介绍了本次实验所使用到的数据集, 此外我们还对样本中数量较少的类别进行了数据增强; 在第4节中, 对基于FPN-ViT模型得到的分类结果进行分析和讨论, 并与其他类似的工作进行了比较. 同时, 我们还对FPN-ViT模型的分类结果进行了可视化分析; 最后, 我们在第5节中对本工作进行了总结.

2 方法

本文提出了一种将FPN引入ViT模型以进行星系图像分类的方法. 传统的ViT模型在处理多尺度图像时存在一定的局限性, 因为它们只能处理固定尺寸的输入图像. 特征金字塔是一种多尺度特征表示的方法, 它通过在不同层级的卷积特征图上应用不同尺度的滤波器来捕捉图像中的局部和全局信息. 我们将FPN与ViT模型相结合, 以提高ViT模型的性能. 其整体结构如图1所示.

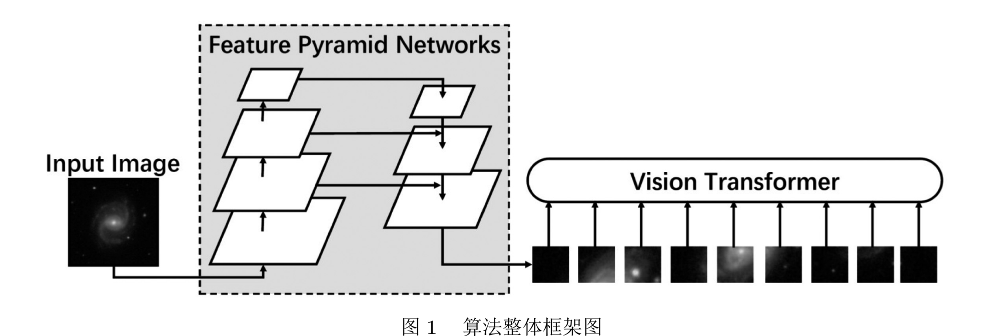


图 1　算法整体框架图

Fig. 1　Overall framework of the algorithm

2.1 特征金字塔网络(FPN)

特征金字塔网络(FPN)是Lin等人在2017年提出的多尺度特征提取器[24]. FPN通过在网络中添加横向连接和上采样操作, 将来自不同层级的特征图进行融合, 构建出金字塔结构的特征表示. 具体来说, FPN在底层特征图上进行上采样操作, 使其与高层特征图的尺寸相匹配, 然后将它们进行逐元素相加, 得到融合后的特征图. 通过这样的操作, FPN在保留高层特征的上下文信息的同时, 还能够有效地利用底层特征的细节信息. FPN的结构如图2所示, 其中Conv2d是二维卷积操作, 1×1表示卷积核大小为1像素高和1像素宽, 3×3则表示卷积核大小为3像素高和3像素宽, s1表示卷积操作的步长(stride)为1, 112、96等数字指的是特征图的尺寸和通道数.

FPN结构的具体计算步骤如下.

(1)输入: 以图像作为输入;

(2)特征提取: 利用CNN进行特征提取, 在CNN中选择适当的层级作为特征提取的起始点, 通常选择较深的层级, 这些层级的特征图具有较大的感受野, 但分辨率较低;

(3)顶层特征: 从特征提取的最深层级开始, 应用一个1×1卷积(或者称为逐点卷积), 生成具有较少通道数的特征图. 这个过程旨在减少计算量, 并为之后的操作做准备;

(4)上采样: 对于比顶层特征分辨率更低的特征图, 通过上采样操作将其尺寸放大, 使其与上一层特征图的尺寸相匹配. 常见的上采样方法包括双线性插值和反卷积;

(5)融合: 将上一步得到的上采样特征图与相邻较浅层的特征图进行逐元素相加(element-wise addition). 这样做可以将细节特征与上下文特征进行融合, 得到多尺度的特征金字塔;

(6)重复: 重复步骤(4)和步骤(5), 直到达到最顶层的特征图;

(7)输出: 最终得到的特征金字塔可以用于目标检测、语义分割等任务. 在目标检测任务中, 通常会使用额外的网络层来预测目标的位置和类别.

输入图像在FPN中的每层特征图及最终输出的特征图像示例如图3所示, 其中Level 1到Level 4对应的是FPN架构不同深度卷积神经网络层所产生的特征图, 最后这些特征图会通过上采样和融合操作生成融合后的特征图(Merged Feature Map).

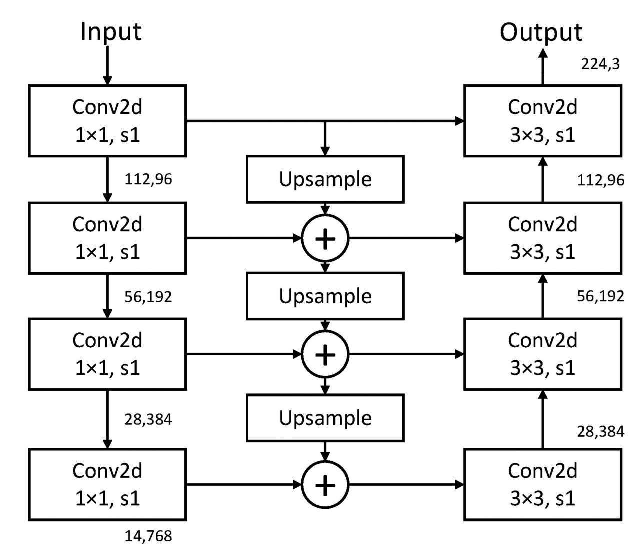


图 2　特征金字塔网络架构图

Fig. 2　Feature pyramid network architecture

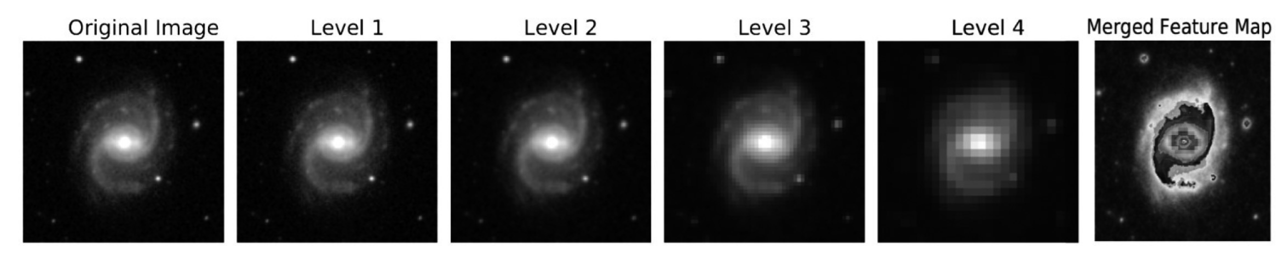


图 3　特征金字塔网络输出特征图像示例图

Fig. 3　Example image of feature pyramid network output feature image

2.2 Vision Transformer

2017年Google的机器翻译团队在神经信息处理系统大会(Conference and Workshop on Neural Information Processing Systems, NIPS)上发表了Attention is all you need的文章, 开创性地提出了在序列转录领域, 完全摒弃了CNN和循环神经网络(Recursive Neural Network, RNN)模型, 只依赖注意力(Attention)结构的简单网络架构, 并命名为Transformer[25]. 2021年, 受Transformer在自然语言处理领域中取得巨大成功的启发, Google团队开发出了一种新的图像分类架构, 并命名为ViT[15], 其基本结构如图4所示.

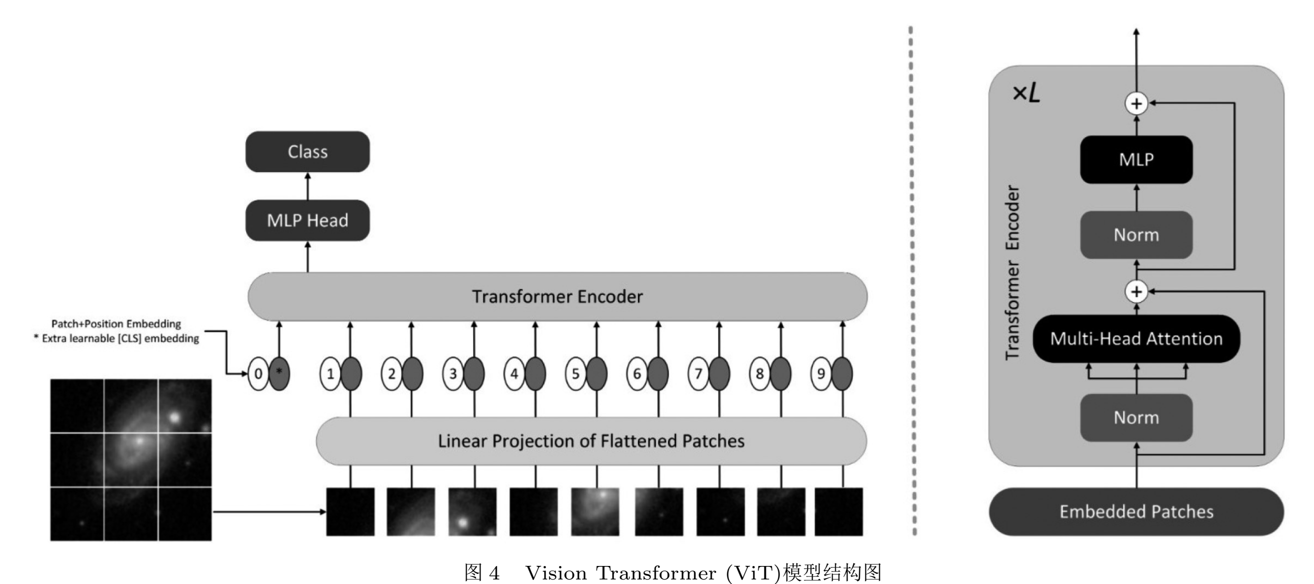


图 4 Vision Transformer (ViT)模型结构图

Fig. 4 Structure of Vision Transformer (ViT) model

ViT模型输入图片首先被切分成固定尺寸的图像块, 之后对展平的图像块进行线性映射(通过矩阵乘法对维度进行变换). 为了保留每个图像块的位置信息, 在图像块送入Transformer编码器之前, 对每个图像块加入了位置编码. 具体计算公式如下:

$$z_0 = [x_{\text{class}}; x_{\text{p}}^1 E; x_{\text{p}}^2 E; \cdots; x_{\text{p}}^N E] + E_{\text{pos}}\,,$$
$$E \in \mathbb{R}^{(P^2\cdot C)\times D}\,,\ E_{\text{pos}} \in \mathbb{R}^{(N+1)\times D}\,, \tag{1}$$

其中, z_0是对整个Transformer Encoder的输入, x_{class}是对输入图像$x \in \mathbb{R}^{H\times W\times C}$ (其中H和W分别是图像的长度和宽度, C是图像的通道数)进行序列化时添加的一个[CLS]标记, 用来累积并包含整个序列的信息. $x_{\text{p}}^i (i = 1\cdots N)$是输入的第$i$个图像块patch, E是线性投影层Linear Projection, E_{pos}是位置编码, P是每个图像块的分辨率, D是图像块的维度, $N = HW/P^2$是输入Transformer编码器的图像块的序列长度.

Transformer编码器由L个标准的Transformer模块组成, 每个模块由层归一化(Layer Normalization, LN)、多头自注意力模块(Multi-head Self-Attention, MSA)、多层感知机(multilayer perceptron, MLP)及残差连接(Residual Connection, R-C)等构成. 具体计算过程如下所示:

$$z_l{}' = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1}\,,\ l = 1\cdots L\,, \tag{2}$$

$$z_l = \text{MLP}(\text{LN}(z_l')) + z_l'\,,\ l = 1\cdots L\,, \tag{3}$$

$$y = \text{LN}(z_L^0)\,, \tag{4}$$

其中, z_l是输入到Transformer Encoder中的图像块序列, l是循环的次数. $z_l{}'$是应用多头自注意力(MSA)模块和残差连接后的中间表示. 对前一层的输出z_{l-1}进行层归一化操作, 然后通过多头自

注意力(MSA)模块, 得到z_l'. z_l是应用多层感知机(MLP)模块和第2个残差连接后的表示. 在得到z_l'后, 它再次经过层归一化, MLP模块和残差连接, 从而得到z_l. z_L^0是L次Transformer Encoder循环结束之后, 最后一层输出的图像块序列z_L中第1个位置上的[CLS]标记, y是Transformer Encoder最终的输出结果.

此外, 该模型需要注意的是只有在大规模数据集上进行预训练再迁移到中小规模数据集的条件下, ViT才能够取得与当时最新卷积结构相媲美的性能.

3 数据

3.1 数据集简介

本文采用的数据集是星系动物园2[26] (Galaxy Zoo 2, GZ2), 该数据集是基于星系动物园项目这个大规模志愿者分类工作而创建的, 并采用了其提供的数据和分类标准. 星系动物园项目源自Kaggle平台上举办的Galaxy Zoo-the Galaxy Challenge比赛, 是一个众包协作的天文学项目, 旨在通过志愿者的分类工作来研究和理解星系的形态和演化.

该比赛的训练集包含了来自斯隆数字化巡天数据SDSS发布的第7个版本数据的61578张带有标签的星系观测图片. SDSS的星系观测数据包括了5个光学波段(u、g、r、i和z), 而在相关研究中常使用前3个波段的数据合成为对应的RGB星系图像. 每张图片的尺寸为$424\times424\times3$像素, 且都有一个1×37的标签向量, 这些标签是根据GZ2志愿者投票分数的修正累计频率值得出的. GZ2对星系的形态进行了11个问题和37个答案的划分. 参考相关工作[9, 14], 在本文中, 我们选择了5类星系数据, 并将其应用于FPN-ViT模型进行分类研究, 包括中间平滑星系(In-between smooth galaxy)、圆形平滑星系(Completely round smooth galaxy)、侧向星系(Edge-on galaxy)、旋涡星系(Spiral galaxy)和雪茄状平滑星系(Cigar-shaped smooth galaxy).

3.2 样本数据选取

对GZ2数据集进行干净样本(well-sampled galaxies)选择需要遵循数据发布白皮书中的规则. 以下是这些规则的描述:

(1)志愿者人数要求—对于每张星系图片, 必须有至少20个志愿者对其进行分类, 这确保了每张图片都得到了足够多的分类意见;

(2)累计投票分数修正值阈值—对于每张图片, 计算得到的累计投票分数修正值必须满足一定的阈值, 该修正值是基于志愿者对该图片的分类结果进行综合得出的;

(3)阈值条件—为了将一张图片分类到特定的星系类别, 必须满足相应的阈值条件. 以旋涡星系为例, 一张图片必须满足该类别的3个阈值条件(一张图片被分类为有特征/盘状结构的频率$f_{\rm features/disk}\geqslant0.430$, 一张图片被分类为非侧向星系的频率$f_{\rm edge-on,no}\geqslant0.715$, 一张图片被分类为旋涡星系的频率$f_{\rm spiral,yes}\geqslant0.619$), 才能被分类为旋涡星系;

(4)平滑星系的特殊情况—由于GZ2的干净样本选择规则较为严格, 对于平滑星系(圆形星系、中间星系和雪茄状星系)3个类别, 可选取的数据样本数目相对较少;

为了确保获得足够数量的样本用于模型训练和测试, 本文适度放宽了平滑星系的阈值选取标准, 将其从0.8降低到了0.5, 而侧向星系和旋涡星系的阈值选取规则仍采用GZ2数据发布白皮书中默认的取值[14].

最终, 按照上述规则, GZ2数据集选择了共计28790张干净样本的星系图片. 这些干净样本经过严格的选取过程, 可以用于进行星系分类和相关研究. 图5为从干净样本原始数据集中随机抽取的5类星系图像示例.

将28790张干净样本按9:1的比例划分为训练集和测试集, 图6给出了每一类星系的训练集和测试集中图像数目, 可以看出5类星系的图片数目满足同分布同比例的要求.

此外, 为了不同类别样本数量之间的平衡性, 本研究将数量较少的雪茄状星系数据通过旋转的方式对训练集的数据进行增强. 我们将雪茄状星系的图像数据分别旋转45°、90°、120°和180°, 旋转后的星系图像示例如图7所示, 其中r表示旋转角度.

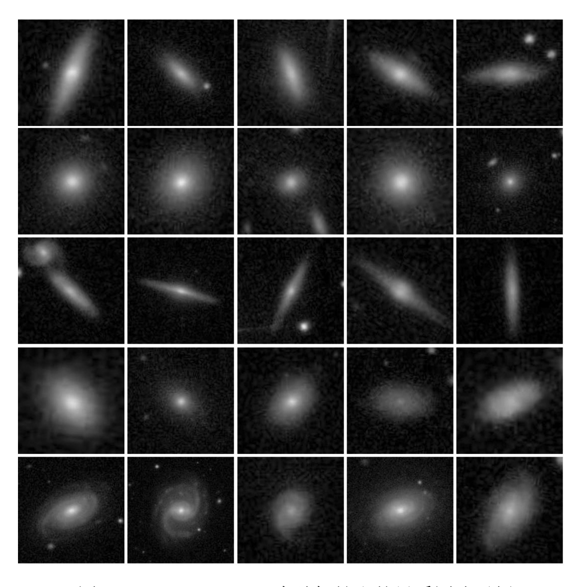


图 5　Galaxy Zoo 2中随机抽取的星系图片示例

Fig. 5　Example images of randomly selected galaxies from Galaxy Zoo 2

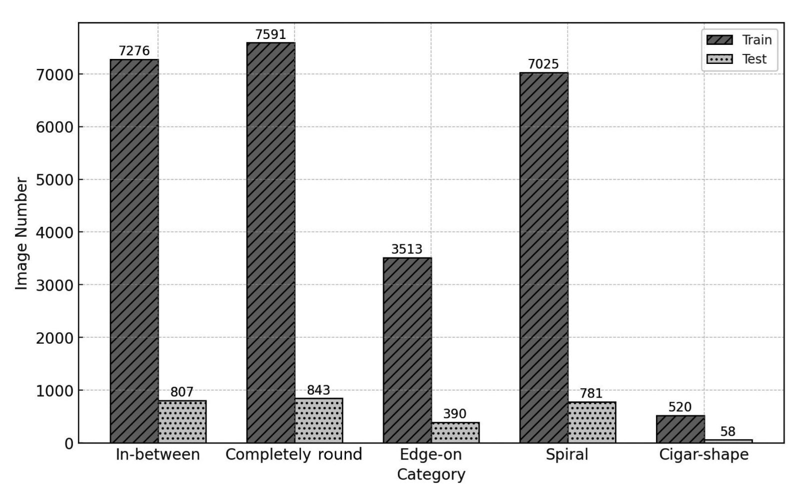


图 6　Galaxy Zoo 2数据分布

Fig. 6　Data distribution of Galaxy Zoo 2

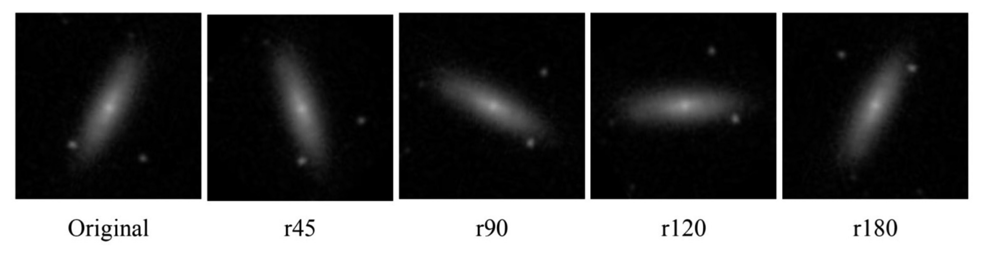


图 7　旋转后的雪茄状星系图像示例

Fig. 7　Example images of a cigar-shaped galaxy after rotation

4 基于FPN-ViT模型的星系形态分类结果

4.1 实验环境

计算基于V100-SXM2-32GB GPU,12 vCPU Intel (R) Xeon (R) Platinum 8255C CPU的服务器进行, 编译器为2021.1版Pycharm-professional, CUDA版本为11.3. 此外, 编程基于pytorch 1.11.0框架采用Python语言实现, 运用了sklearn、Scikit-image、transforms等python库.

4.2 结果分析

为了验证FPN-ViT模型的分类性能, 本工作基于基础的FPN-ViT B/16模型, 并采用准确率、精确率、召回率以及F1分数等评价指标来衡量模型的分类性能. 表1给出了FPN-ViT在各类星系中的最好分类效果. 可以看出除雪茄状星系外, 每个类别星系的分类准确率超过了98%, 并且精确率、召回率以及F1分数也都在97%以上, 而雪茄状星系的各项分类评价指标均未超过90%, 但是也都在82%以上. 雪茄状星系的分类效果不佳主要是由于其数据量过少造成的. 同时, 在5类星系分类的平均情况下, 准确率为95.2%、精确率为95.2%、召回率为95.0%、F1分数为95.2%, 验证了FPN-ViT模型对星系的形态分类有着很好的鲁棒性.

表 1 基于FPN-ViT模型5类星系中的分类性能

Table 1 Classification performance for 5 classes of galaxies based on the FPN-ViT model

Class	Galaxy	Accuracy	Precision	Recall	F1 Score
0	In-between	98.7%	98.9%	98.5%	98.8%
1	Completely round	98.4%	98.3%	98.0%	98.1%
2	Edge-on	98.0%	98.0%	97.6%	97.9%
3	Spiral	98.3%	98.1%	98.0%	98.1%
4	Cigar-shape	82.6%	82.9%	82.8%	83.1%
	Average	**95.2%**	**95.2%**	**95.0%**	**95.2%**

与此同时, 图8采用接收者操作特征(Receiver Operating Characteristic, ROC)曲线并计算出ROC曲线下面积(Area Under the Curve, AUC)来评估模型性能. 结果表明模型对每个类别的星系都有较好的分型效果, 除数据样本较少的雪茄状星系的AUC值为0.975外, 每个类别星系的AUC值均在0.98以上. 图8中, area表示ROC曲线下的面积, 即具体的AUC值.

图9中展示了在GZ2数据集上通过FPN-ViT模型进行测试的混淆矩阵. 混淆矩阵表示的是一个多类别分类模型的性能评估结果, 在该矩阵中, 0–4分别代表不同形态的星系图像. 结果显示, 中间星系和圆形星系的分类准确率较高, 而侧向星系、旋涡星系和雪茄状星系的分类存在一些错误, 特别是雪茄状星系的分类表现较差. 雪茄状星系中有7个被错误地归类为中间星系, 12个被错误地归类为侧向星系, 还有3个被错误地归类为旋涡星系, 这可能是因为雪茄状星系的数据量较少, 模型在训练过程中可能没有充分学习到它们的形态特征, 导致分类结果不佳.

FPN是为了解决图像识别过程中由于图片大小差异导致的识别困难而提出的算法, 为了验证FPN结构的有效性, 我们对GZ2数据集中的星系图像进行了不同程度的缩放. 我们通过调整scale参数来控制图像的缩放比例, 在实验中分别设置了0.35、0.5、0.65、0.8的scale值. 缩放后的星系图像

如图10所示. 对于调整大小之后的星系图像, 我们使用FPN-ViT模型对其进行分类, 结果如表2所示. 实验结果显示, 调整后的GZ2数据集使用FPN-ViT模型对不同形态的星系图像进行分类的平均准确率均在90%左右, 说明FPN-ViT模型对于不同大小和分辨率的星系图像都有着较好的分类效果, 验证了在ViT模型中引入FPN网络结构的有效性.

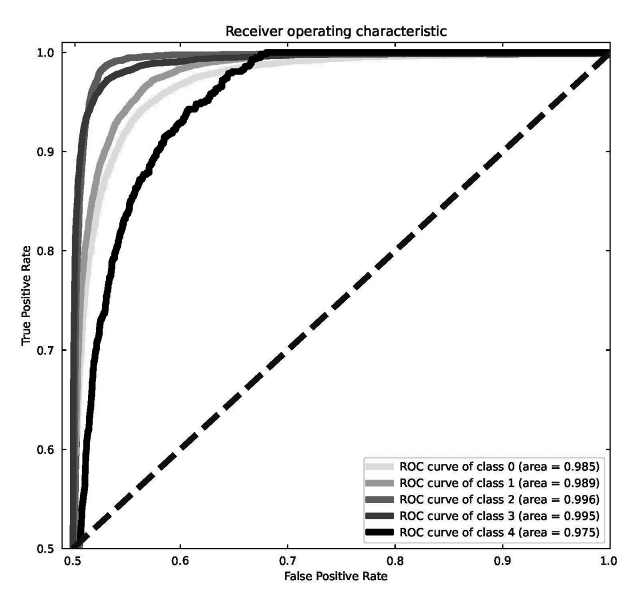


图 8　FPN-ViT在不同形态的星系分类中的ROC曲线

Fig. 8　ROC curve of the FPN-ViT model in the classification of galaxies of different morphologies

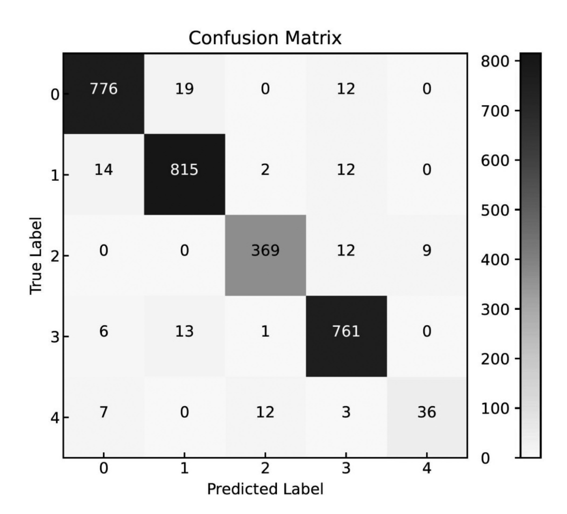


图 9　FPN-ViT对5类星系分类的混淆矩阵

Fig. 9　Confusion matrix for classification of five types of galaxies using FPN-ViT

表 2　不同大小星系图像的分类结果对比

Table 2　Comparison of classification results for galaxy images of different sizes

Model	Different scales	Accuracy
FPN-ViT	scales=0.35	90.2%
	scales=0.50	93.6%
	scales=0.65	94.3%
	scales=0.80	94.7%
	scales=1.0	95.2%

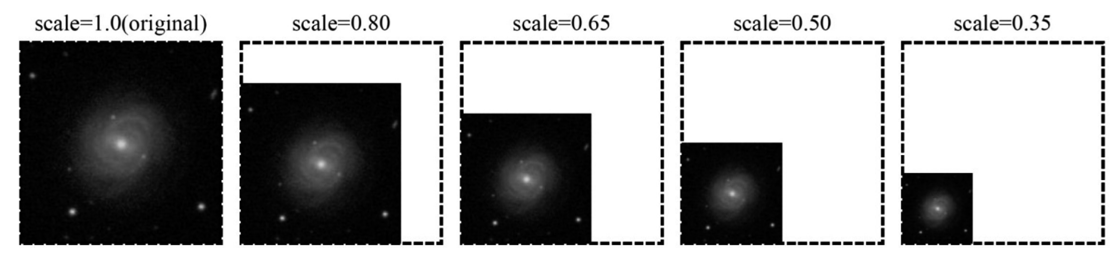


图 10　不同缩放程度的星系图像示例

Fig. 10　Examples of galaxy images at different scale levels

此外, 我们往原始星系图像上添加了不同程度的高斯噪声和椒盐噪声, 以此来验证FPN-ViT模型对低信噪比星系图像的泛化能力. 本文通过调节高斯分布标准差(sigma)的大小来控制添加高斯噪声程度, 在实验中设置了分别为5、15、25的sigma值. 椒盐噪声就是给图片添加黑白噪点, 通过设置amount来控制添加噪声的比例, 实验中设置了amount分别为0.05、0.1、0.2. Sigma和amount的

值越大添加的噪声越多, 图像损坏更加严重. 添加噪声后的星系图像如图11所示, 基于FPN-ViT模型在加入噪声的星系形态中分类结果如表3所示. 添加噪声后的星系图像与未加噪声的星系图像相比, 从分类性能的角度来看, FPN-ViT模型的整体分类效果有所下降. 然而, 在噪声影响下, 模型的整体分类精度仍能保持在70%以上, 这表明FPN-ViT模型在处理低信噪比星系图像的分类任务上表现稳定, 并展现出相当不错的泛化能力.

在星系形态分类研究中, 星系图像亮度的变化与观测的距离密切相关. 由于星系在不同距离下观测到的亮度会发生变化, 因此使用不同亮度的图像可以模拟不同距离下的星系观测, 并研究形态分类算法对于距离效应的鲁棒性. 在实验中, 我们分别设置了0.5、0.75、1.5、2.0的亮度值(brightness)来调整GZ2数据集中星系图像的亮度, 调整后的图像如图12所示. 对于不同亮度的星系图像使用FPN-ViT模型进行分类的结果如表4所示, 实验验证了FPN-ViT模型对不同亮度的星系图像分类有着较好的鲁棒性.

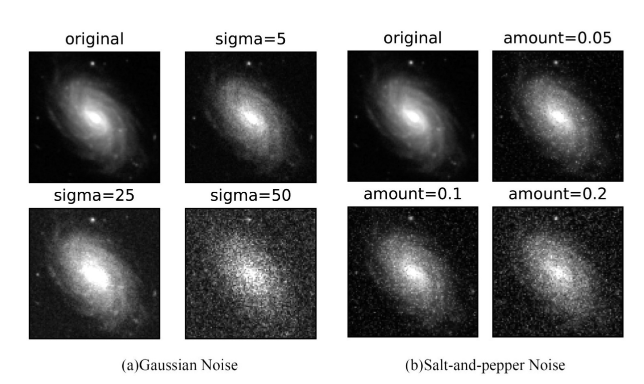


图 11 添加不同类型和程度噪声的星系图像示例

Fig. 11 Examples of galaxy images with different types and levels of noise

表 3 添加噪声后星系图像的分类结果对比

Table 3 Comparison of galaxy image classification results after adding noise

Model	Add noise type	Add noise level	Accuracy
FPN-ViT	Gaussian noise	sigma=0	95.2%
	Gaussian noise	sigma=5	91.1%
	Gaussian noise	sigma=25	82.4%
	Gaussian noise	sigma=50	73.7%
	Salt-and-pepper noise	amount=0	95.2%
	Salt-and-pepper noise	amount=0.05	91.3%
	Salt-and-pepper noise	amount=0.1	80.1%
	Salt-and-pepper noise	amount=0.2	71.5%

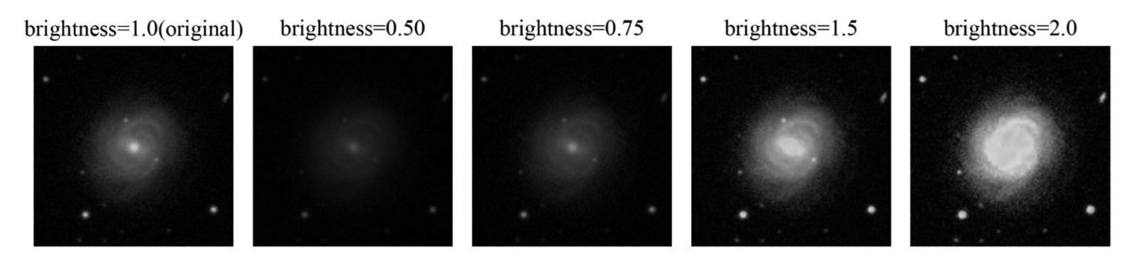


图 12 不同亮度的星系图像示例

Fig. 12 Example of galaxy images with different brightness

表 4 不同亮度星系图像的分类结果对比

Table 4 Comparison of classification results for galaxy images of different brightness

Model	Different brightness	Accuracy
FPN-ViT	brightness=0.50	89.3%
	brightness=0.75	95.3%
	brightness=1.50	94.7%
	brightness=2.00	93.1%
	brightness=1.00	95.2%

然而在实际的天文观测中, 观测的距离与拍摄图像大小、亮度和信噪比之间存在一些关系. 这些关系可以归结为以下几点.

(1)图像大小: 一般来说, 观测距离越远, 星系或天体在图像上所占的角尺寸越小. 这是由于观测距离的增加导致星系或天体的视角缩小. 因此, 随着观测距离的增加, 图像中的天体大小会变小;

(2)亮度: 观测距离与拍摄图像中的天体亮度之间的关系取决于天体的固有亮度和观测条件. 一般来说, 观测距离增加时, 天体的亮度会减弱. 这是由于星系或天体的辐射能量在传播过程中的衰减效应. 因此, 随着观测距离的增加, 拍摄图像中的天体亮度会降低;

(3)信噪比: 观测距离对信噪比也有影响. 观测距离越远, 图像中的天体所接收到的光信号可能会减弱, 同时噪声的影响也会更加显著. 这会导致图像的信噪比降低, 即图像中的信号相对于噪声的强度减弱. 因此, 观测距离的增加会对图像的信噪比产生负面影响.

因此, 准确地描述观测距离与图像大小、亮度和信噪比之间的关系需要综合考虑多个因素. 于是我们扩展了实验, 通过距离参数(distance值)来模拟真实观测中的观测距离, 使得观测距离越远, 拍摄得到的星系图像大小越小, 同时亮度越暗、信噪比越低. 在实验中, 我们分别设置了0.5、0.75、1.5、2.0的distance值来模拟观测距离, 设置后的图像如图13所示. 在模拟的不同观测距离下, 我们计算出该距离下星系图像的亮度和峰值信噪比(Peak Signal-to-Noise Ratio, PSNR), 峰值信噪比是通过比较原始图像和受噪声影响图像之间的均方根误差来计算噪声对图像质量的影响. 所以, 峰值信噪比越大表示图像的信号(原始图像)与噪声(噪声图像)之间的比值越大, 也就是原始图像与噪声图像之间的差异越小, 因此噪声越小. 对于模拟的不同观测距离的星系图像, 我们使用FPN-ViT模型进行5分类实验结果如表5所示. 实验结果表明, FPN-ViT模型对于模拟的不同观测距离下的星系图像分类准确率均在75%以上, 有着较好的鲁棒性.

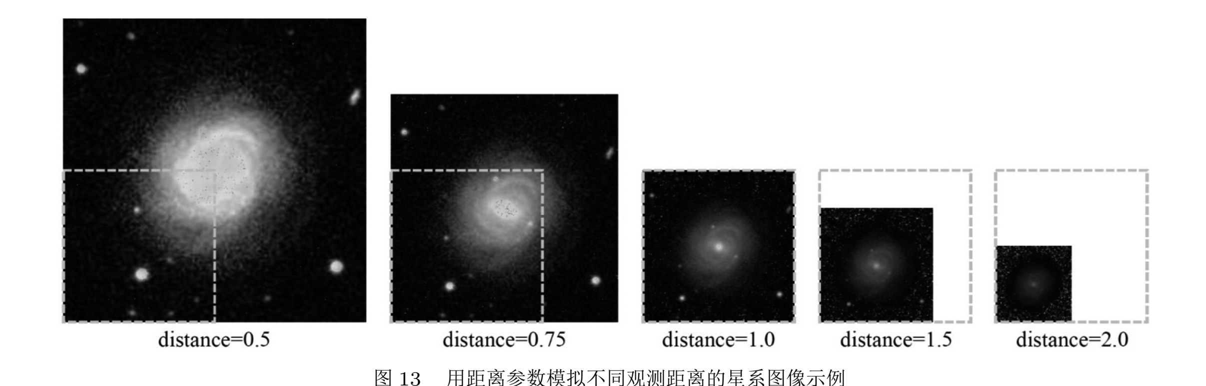


图 13 用距离参数模拟不同观测距离的星系图像示例

Fig. 13 Examples of galaxy images simulated with distance parameters for different observation distances

我们将FPN-ViT模型获得的分类效果与传统的Vision Transformer模型进行对比, 这两项工作所用的数据来源相同, 且训练集和测试集划分比例一致. 具体分类结果对比情况如表6所示, 基于Transformer架构的两种模型在星系系统分类方面都取得了较高的准确率. 并且我们对比了基础的

ViT模型和改进后的FPN-ViT模型, 发现FPN-ViT模型的各个评价指标对比ViT模型都在一定程度上有所提高. 这说明FPN-ViT模型相比于基础的ViT模型, 在星系的形态分类任务中有较好的效果.

表 5 不同观测距离下星系图像的各项数值
Table 5 Values of galaxy images at different observing distances

Distance	Brightness	PSNR	Image Size	Accuracy
0.50	49.06	42.37	424 × 424	94.9%
0.75	38.42	38.25	318 × 318	95.1%
1.00	28.09	35.33	212 × 212	94.6%
1.50	26.70	33.39	159 × 159	82.0%
2.00	25.79	32.40	106 × 106	75.2%

表 6 FPN-ViT模型与ViT B/16对GZ2的分类结果对比
Table 6 Comparison of classification results of FPN-ViT model and ViT B/16 for GZ2

Model	Accuracy	Precision	Recall	F1-Score
ViT B/16[15]	94.6%	94.1%	94.2%	94.1%
FPN-ViT (this work)	95.2%	95.2%	95.0%	95.2%

4.3 分类结果的可视化

为了从分类结果探索星系形态特征的信息, 我们将测试集的分类结果进行可视化. 在这个部分, 我们使用的是t-SNE算法对FPN-ViT模型的分类结果进行可视化分析. t-SNE算法是一种用于多维数据缩放的非线性降维算法[27], 它可以保留数据样本数据的局部结构, 并获得与原始高维度数据相似度更高的低维度数据. 由于其在高维数据缩放到较低维数据方面具有显著的效果, 因此在机器学习中应用广泛. t-SNE算法将数据点之间的相似性转化为概率, 原始高维空间中的相似性用高斯分布表示. 嵌入空间的概率用T分布表示, 从而将高维空间的数据映射到低维空间并进行可视化表示.

图14是FPN-ViT模型对星系形态分类结果的可视化. 从图中可以看出, 各类星系的簇都有着较为清晰明确的界限, 这表明了FPN-ViT模型对于星系形态分类的效果较好. 而侧向星系和雪茄状星系的边界有极小部分相连在了一起, 是因为这两类星系数据样本较少, 且形状较为相似, 所以导致了这两类星系的部分图像被错误分类.

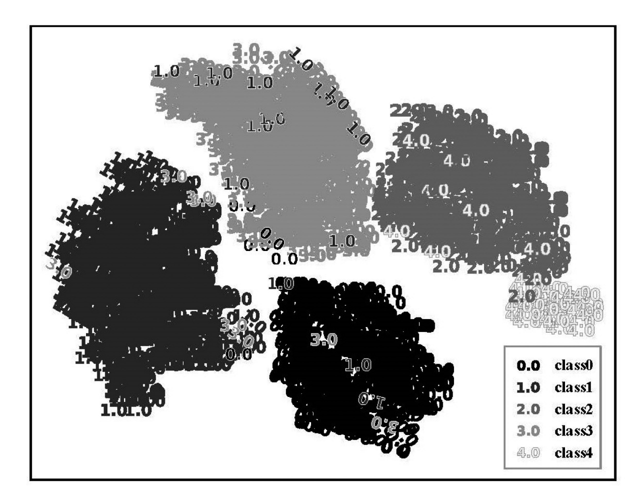


图 14 FPN-ViT分类结果可视化

Fig. 14 Visualization of FPN-ViT classification results

5 总结和展望

随着巡天项目的不断深入、天文观测范围的扩大和观测技术的提升, 巡天项目产生的天文数据规模不断增大, 传统的数据处理方法将无法满足大规模数据的处理需求. 本文鉴于深度学习在天文数据中广泛应用和Transformer方法在NLP领域获得的巨大成功, 将FPN-ViT模型应用于星系形态的分类研究.

其中基于Transformer的分类模型在星系形态分类上都获得了较高的准确率, 在FPN-ViT模型中整体平均准确率为95.2%, 平均精确率为95.2%、平均召回率为95.0%、平均F1分数为95.2%. 相对于基于CNN的分类模型有了一定程度的提升, 证明了基于Transformer的分类模型可以应用于星系的形态分类中. 同时, FPN-ViT对于低信噪比星系图像的分类准确率均在70%以上, 说明该模型对于低信噪比星系图像也有着较好的泛化能力. 此外, 本工作中还使用t-SNE算法对模型的分类结果进行可视化, 可以更加直观地看出FPN-ViT模型对于星系形态分类的效果.

在未来, 中国空间站望远镜(China Space Station Telescope, CSST)和大型综合巡天望远镜(Large Synoptic Survey Telescope, LSST)等大型望远镜计划在几年内发射, 它们将为天文学研究提供更广阔的观测范围和更详细的天文数据. 本文采用的FPN-ViT模型对后续数据分析提供了更多可能, 这意味着该模型可以应用于更广泛的天文数据集, 不仅局限于本文提到的GZ2数据集. 我们将会继续对FPN-ViT模型进行探索和研究, 并将用该模型对非本文所述形态的星系图像进行分类研究, 同时还将会重点研究FPN-ViT模型中网络结构对形态分类效果的影响, 进一步验证该模型在星系形态分类中的有效性.

参考文献

[1] Hubble E P. ApJ, 1926, 64: 321
[2] Dressler A. ApJ, 1980, 236: 351
[3] Blanton M R, Moustakas J. ARA&A, 2009, 47: 159
[4] Sparke L S, Gallagher J S. Galaxies in the Universe: An Introduction. Cambridge, UK: Cambridge University Press, 2007: 278-314
[5] Lintott C J, Schawinski K, Slosar A, et al. MNRAS, 2008, 389: 1179
[6] York D G, Adelman J, Anderson Jr J E, et al. AJ, 2000, 120: 1579
[7] Cui X Q, Zhao Y H, Chu Y Q, et al. RAA, 2012, 12: 1197
[8] Gardner J P, Mather J C, Abbott R, et al. PASP, 2023, 135: 068001
[9] Zhu X P, Dai J M, Bian C J, et al. Ap&SS, 2019, 364: 55
[10] 艾霖嫔, 徐权峰, 杜利婷, 等. 天文学报, 2022, 63: 44
[11] Wei S, Li Y, Lu W, et al. PASP, 2022, 134: 114508
[12] He Y, Zhang Y, Chen S, et al. 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Chongqing: IEEE, 2023, 6: 1648
[13] Hui W, Jia Z R, Li H, et al. Journal of Physics: Conference Series, 2022, 2402: 012009
[14] Li G, Xu T, Li L, et al. MNRAS, 2023, 523: 488
[15] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations (ICLR), 2021
[16] Gheflati B, Rivaz H. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Glasgow: IEEE, 2022: 480
[17] Gao X, Khan M H M, Hui R, et al. 2022 3rd International Conference on Next Generation Computing Applications (NextComp). Mauritius: IEEE, 2022: 1
[18] Tanzi L, Audisio A, Cirrincione G, et al. Injury, 2022, 53: 2625-2634
[19] Chu X, Tian Z, Zhang B, et al. Conditional Positional Encodings for Vision Transformers. International Conference on Learning Representations (ICLR), 2023
[20] Han K, Xiao A, Wu E, et al. Advances in Neural Information Processing Systems, 2021, 34: 15908
[21] Yuan L, Chen Y, Wang T, et al. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 558
[22] Wang W, Xie E, Li X, et al. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 568
[23] Wu H, Xiao B, Codella N, et al. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 22
[24] Lin T Y, Dollár P, Girshick R, et al. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos: IEEE, 2017: 936
[25] Vaswani A, Shazeer N, Parmar N, et al. Advances in Neural Information Processing Systems. Long Beach: Curran Associates Inc, 2017, 30: 6000
[26] Willett K W, Lintott C J, Bamford S P, et al. MNRAS, 2013, 435: 2835-2860
[27] Devassy B M, George S. Forensic Science International, 2020, 311: 110194

Classification of Galaxy Morphology Based on FPN-ViT Model

CAO Jie[1] XU Ting-ting[1] DENG Yu-he[1] LI Guang-ping[1] GAO Xian-jun[1] YANG Ming-cun[1] LIU Zhi-jing[1] ZHOU Wei-hong[1,2]

(1 School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650504)
(2 Key Laboratory of the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, Kunming 650011)

Abstract With the development of artificial intelligence technology, the research of galaxy morphology classification using deep learning methods has made great progress, but there are still shortcomings in classification accuracy, automation and spatial characteristics representation of galaxies. The Vision Transformer model has good robustness in galaxy morphology classification, but has limitations in handling multi-scale images. In this paper, we propose to introduce the Feature Pyramid Networks (FPN) into the Vision Transformer (ViT) model to classify galaxies. The results show that the average accuracy, precision, recall, and F1-score of the FPN-ViT model are above 95%, and the indexes are improved compared with the traditional ViT model. Meanwhile, we add different levels of Gaussian noise and pretzel noise to the original galaxy images to verify that the FPN-ViT model can obtain better classification performance for low signal-to-noise ratio data. In addition, to evaluate the model comprehensively, the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm is used to visualize and analyze the classification results, which can show the effect of FPN-ViT model on galaxy morphology classification more directly. The application of FPN network to the classification of galaxy morphology by ViT model is a new attempt, which is of great importance for the subsequent research.

Key words methods: data analysis, techniques: image processing, galaxy: general